THE NOTES TO METHODS FOR INDUSTRIAL CLUSTER IDENTIFICATION[1]

Jan Stejskal, Petr Hájek
  Univerzita Pardubice, Fakulta ekonomicko-správní

Abstract: *The conception of the regional policy and regional development is still an evolving process. The economic growth is the limelight in context of the European regional policy. The main growth determinant is also increasing competition of the municipalities, regions and countries. It can be realized thanks to engines (centres of excellence, clusters) support. These centres are the carriers of the main part of the local and regional economic development. At these levels there are many tools of local and regional development used.*

Key words: *business climate, industrial clusters, Czech Republic, regional policy*

1. Easy Methods of Industrial Clusters Identification

Many regional scientists [1,2,3,4] patronize the ideas of diversification the identification methods according to the level where the clusters are analyzed. National (macro-level), industrial branch (mezzo-level) and firm level (micro-level) can be recognized.

The mezzo-level was considered as the most important in the 90-ties of the 20th century [1]. The clusters are identified in European countries and regions at all levels contemporary [5]. We recommend to diversify also national (state or country), regional (NUTS II or NUTS III) and also local level (municipal or micro-regional). This concept seems to be more suitable in the practice of small European regions. There are many methods applied for cluster identification.

The presented methods of potential clusters identification can be used at macro-level and regional level. The clusters can not be identified in locations with lack of networking and trust. These two premises are the basic conditions of successful clusters. Firms without these conditions are called "conflux" of the firms, no cluster. The better concept than conflux is known as a quasi-cluster.

Following methods for cluster identification were used in various studies [6,7,8,9,10] so far:

a) Expert examination - it must be based on many detailed information, but does not make a generalization possible,

b) Indexes of specialization (localization coefficient, LQ) – they are frequently used due to easy calculation; these indexes are supplementary only; they are oriented on specific fields; thereupon the result of this method is not fully relevant for cluster formation decision). Localization coefficients compare the characteristics of branches (number of employees, sales and added value) at the regional and national level. The results of the LQ show the dominant localization of enterprises in the given branches. The localization quotients for the number of employees is defined as follows

$$LQ_i = \frac{z_i / z}{Z_i / Z} ,$$
(1)

where: - LQ_i is localization coefficient of the i-th branch (employees),
   - z_i is the number of employees of the i-th branch in a region,

- z is the total number of employees in the region,
- Z_i is the number of employees of the i-th branch in the Czech Republic,
- Z is the total number of employees in the Czech Republic.

$$LQ_i^v = \frac{v_i/v}{V_i/V},$$
(2)

where: - LQ_i^v is localization coefficient of the i-th branch (turnover, value added)
- v_i is the value of output (turnover, value added) of the i-th branch in a region,
- v is the value of output in the region,
- V_i is the value of output (turnover, value added) of the i-th branch in the Czech Republic,
- V is the value of output in the Czech Republic.

c) Input-output analysis of business relationships (IOA) - this method can identify the relationships among firms which are necessary for cluster initiatives; the drawbacks of this method are quick obsolescence, low accuracy and the inability of its application in small regions,

d) Input-output analysis of innovations - this method is also known as index of innovative activity; it does not focus on the clusters actually,

e) Network analysis and graph theory - this method is applied as a visualization tool,

f) Statistical and economic compendium - it is actual, but costly; the objectivity of the results interpretation is crucial,

g) Comparative advantages analysis – ratio of regional amount of export in every sector to export of all sectors, this ration must be compared with ratio from compared countries.

The cluster potential must be integrated with development of the competitiveness advantage which was defined by Porter in his diamond, Porter Diamond (PD). Generally this potential conception can be identified with:

- development ability in future,

- ability of "move" local (regional or national) economy,

- space making for the new innovations birth,

- creating some vacancies.

The cluster potential is the useful tool for rating of industrial branches and new trends identifying as the tool for regional or spatial management. This concept can be used also in businesses, industrial branches and regions.

Based on presented reasons we suggest a statistical method based on Porter diamond for cluster identification. The objective of our method is the analysis of empirical data by multivariate statistical methods. The data are represented by different competitive advantages [2]. The results can be interpreted for industrial branches. This way the branches with the biggest cluster potential can be identified.

2. Multivariate Statistical Methods for Industrial Clusters Identification

The main applications of principal component analysis (PCA) and factor analysis (FA) are the reduction of the number of variables and the detection of the structure in the relations between variables. It is possible to realize the following advantages by the PCA and FA: more effective insights, the reduction of noise in the data and dimension reduction. Therefore, the

result of the PCA and FA is represented by a smaller set of variables that explain most of the variance in the original data, in more compact and insightful form.

The PCA represents a procedure that transforms a number of correlated variables $x_1, x_2, \ldots, x_n$ into a smaller number of uncorrelated variables $C_1, C_2, \ldots, C_u$ called principal components. The principle components $C_1, C_2, \ldots, C_u$ represent the linear combinations of the original variables $x_1, x_2, \ldots, x_n$. The principle of this method lies in the fact that a variable with a higher variation explains a higher proportion of the variation in the dependent variable compared to a variable with a lesser variation. So, the original set of variables $x_1, x_2, \ldots, x_n$ is transformed into the set of variables $C_1, C_2, \ldots, C_u$, where u<n. The variables $C_1, C_2, \ldots, C_u$ are uncorrelated and represent the most of the original variation.

The PCA model is defined as follows

$$\mathbf{X} = \mathbf{T} \times \mathbf{P}^T + \mathbf{E}, \tag{3}$$

where: $\mathbf{X}$ is the original data matrix,

$\mathbf{T}$ is the component score matrix,

$\mathbf{P}^T$ is the transposed component loadings matrix,

$\mathbf{E}$ is the residual matrix which represents the unexplained component of the model.

Let $T = (t_1, t_2, \ldots, t_A)$, $P^T = (p_1^T, p_2^T, \ldots, p_A^T)$, and u is the number of the principal components, the equation (1) can be extended this way

$$\mathbf{X} = t_1 \times p_1^T + t_2 \times p_2^T + \ldots + t_u \times p_u^T + \mathbf{E}. \tag{4}$$

The FA is applied to reveal the dependencies and relations in the data matrix $\mathbf{X}$ structure. The analysis makes it possible to find factors $F_1, F_2, \ldots, F_v$ that represent the original set of variables. Summarization and data reduction are the favourable results of the FA application. Differences between FA and PCA arise from the fact that they are based on different models. The FA can be used as the user is interested in making statements about the factors that are responsible for the observed responses, while the PCA is mostly used as a data reduction method. In the PCA all variability in an item should be used in the analysis, while in principal factors analysis we only use the variability in an item that it has in common with the other items.

3. Example of Statistical Methods Usage in Practical Case

The assignment of a cluster potential to sample of 72 enterprises of Pardubice region is the goal of the modelling. The enterprises are divided into the following branches.

Table 1 The Branches of the monitored enterprises

FoI – Food industry	MaI – Machinery industry
ChI – Chemical industry	FuI – Furniture industry
TI – Textile industry	EI –Electrical industry
MeI – Metalworking industry	BI – Building industry

Localization coefficients are held as a popular method in industrial clusters potential identification. Three alternatives of localization coefficients are used in the evaluation of the firms in the Pardubice region, Fig. 1. The influence of the number of employees, turnover and value added. The values of all localization coefficients are above average for machinery and electrical industry. It means there could be found a specialization in these industrial branches which is bigger than in the rest of regions of the Czech Republic. Localization coefficient LQ_e gives a bit higher values than the localization coefficients LQ_t and LQ_{va}. A multivariable

localization coefficient can not be made as a mean value without taking into accounts the weights and specifics of variables.
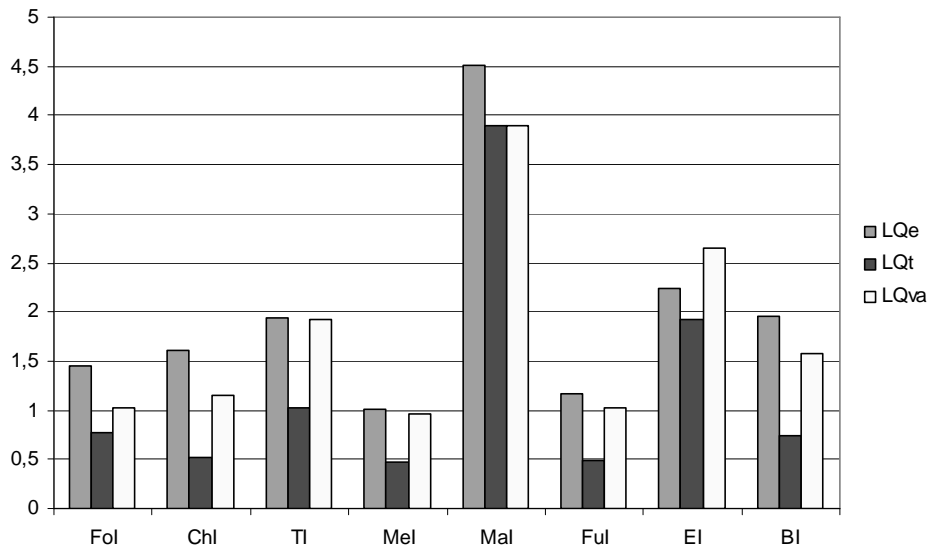

Fig. 1 Industrial clusters potential measured by the localization coefficients LQ_e, LQ_t and LQ_{va}.

The localization coefficient LQ_i depends only on one variable which has not to be crucial in the industrial cluster identification in a specific branch (human resources are not necessary in a capital demanding industry branch, etc.). It was primary intended to help to enterprises in their localization decision-making process. It is suitable to use more variables to achieve a more efficient decision.

Based on the presented facts, a methodology enabling the use of their advantages is suitable to realize. The designed methodology is based on the field research data and makes the calculation of industrial cluster potential possible. The calculation is realized by the statistical methods (PCA and FA) reducing the input space dimension. The input data was acquired by a field research. The enterprises answered questions concerning four principal areas (human resources availability, infrastructure, capital and natural resources availability). The enterprises assigned subjective ratings (from 1 to 4) to these factors, where 1 is a fully satisfaction and 4 is a fully dissatisfaction. The missing values are replaced by the means of variables. The data was pre-processed by means of standardization as the dependency on units was removed. The following variables are considered as the input variables:

x_1 - human resources availability,
x_2 - capital resources availability,
x_3 - infrastructure availability,
x_4 - natural resources availability,
x_5 - growth of employment,
x_6 - growth of turnover.

The goal of the PCA modelling is the reduction of 6 input variables to less number of variables which represent most of the variation. The results of the experiments show that it is suitable to create two principle components. The first principle component represents 33.04% and the second principle component represents 17.47% of input data variation (Fig. 2). Further, its eigenvalue is higher than the given threshold value 1.
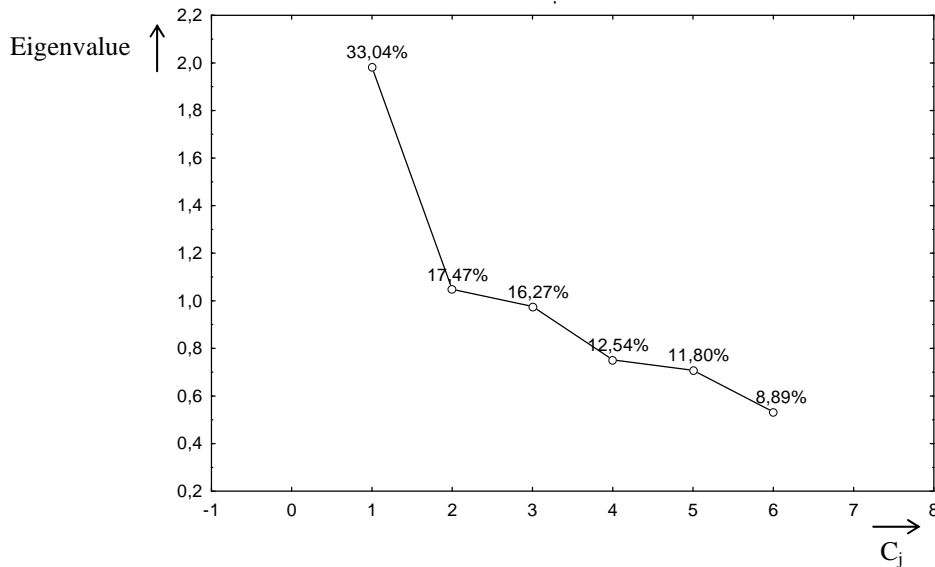
Fig. 2 The components and their eigenvalues

The loadings (weights) of input variables $x_1, x_2, \ldots, x_6$ in the first and second principal components are presented in Fig. 3. According to this, the first principle component represents the variables x_3, x_4, x_5 and x_6 (resources availability) and the second principle component represents the variables x_1 and x_2 (growth of employment and turnover).
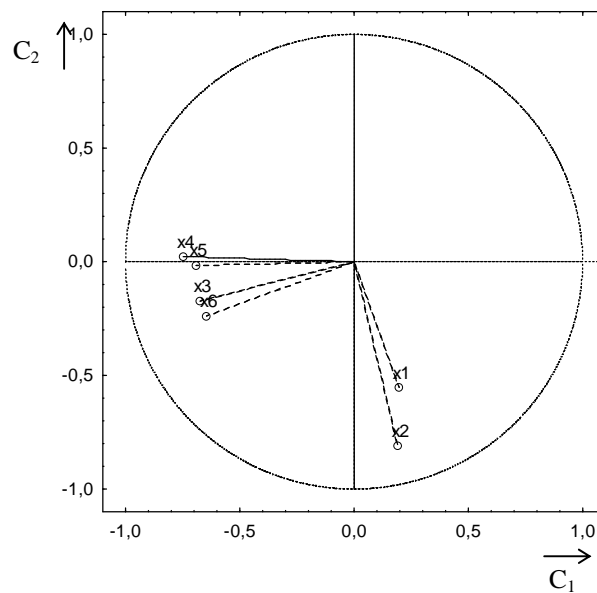

Fig. 3 The loadings of the variables $x_1, x_2, \ldots, x_6$ in the components C_1 and C_2.

Based on these loadings, the values of the first component C_1 (cluster potentials CP_{PCA}) for the individual industry branches can be calculated this way

$$CP_{PCA} = 0{,}197 \times x_1 + 0{,}192 \times x_2 - 0{,}674 \times x_3 - 0{,}747 \times x_4 - 0.691 \times x_5 - 0.646 \times x_6. \qquad (5)$$

The goal of the FA modelling is to reveal the dependencies in the data matrix. The input data is the same as in the PCA modelling. The results of the modelling show that the eigenvalue of the factor 1 (2.23) is higher than its communality (1.00). The loadings of the input variables for the factor 1 are presented in Table 2.

As presented in Table x, factor F1 accounts for 25 percent of the variance and factor F2 for 8.8 percent. The variances extracted by the factors are called the eigenvalues. The sum of the eigenvalues is equal to the number of variables. First, only factors with eigenvalues greater

than 1 can be retained.

Table 2 Factors and their eigenvalues

Factor	Eigenvalue	Variance	Cumulative eigenvalue	Cumulative variance
F1	1,501	25,015	1,501	25,016
F2	0,528	8,807	2,003	33,823

The Porter diamond variables are highly correlated amongst themselves, and the input-output analysis variables are highly intercorrelated amongst themselves. The correlations across these two types of variables are comparatively small. It seems that there is one relatively independent factor F1.

The factor loadings are shown in the Fig. 4. Each variable is represented as a point. In this figure, the axes can be rotated in any direction without changing the *relative* locations of the points to each other; however, the actual coordinates of the points, that is, the factor loadings would of course change. There are various rotational strategies that have been proposed. The goal of all of these strategies is to obtain a clear pattern of loadings, that is, factors that are somehow clearly marked by high loadings for some variables and low loadings for others. Typical rotational strategy is varimax (variance maximizing*)*, as presented in Fig. 4. The factor F1 (F2) is mostly represented with the variable x_4 (x_3). Variables x_5 and x_6 have similar loadings in both factors, while variables x_1 and x_2 are completely irrelevant. The factor scores of factor F1 for each industrial branch is presented in Fig. 5
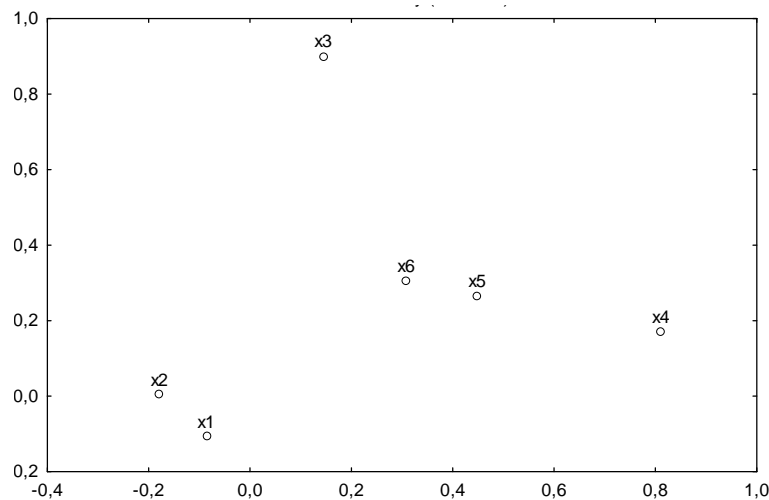

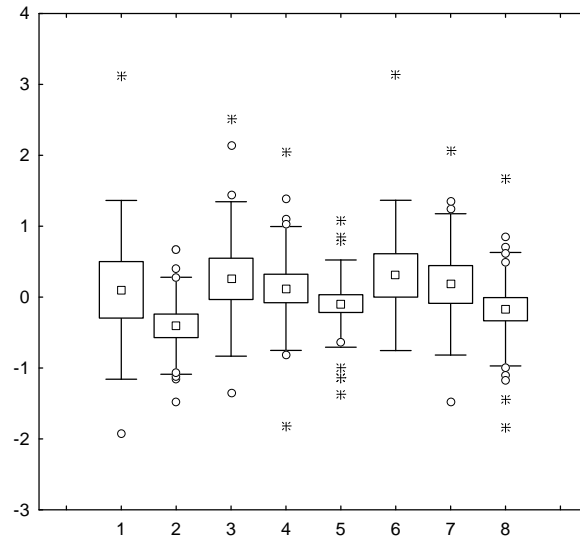Fig. 4 Factor loadings obtained after varimax rotational strategy

Fig. 5 Factor scores for all industrial branches

4. Conclusion

The methods used in this article are showed with some fundamental characteristics in table Y. The use of only one method of analysis is not enough for cluster potential identification. One method can not describe economic reality in which the clusters birth and exist.

Table 3: The characteristics of methods for cluster potential analysis

	LQ	IOA	PD	PCA
Data	Statistics and databases	Statistics and databases	Empirical data	Preprocessed empirical data
Number of factors	Only 1, but variable	2	4	2 and more
Factors	Employment, Turnover, or Value added.	Employment and Turnover.	Demand, Sources, Industrial branches, Strategy.	Demand, Sources, Industrial branches, Strategy, Employment, and Turnover.
Predicative ability	Low (only orientation information)	Low (only additional information)	Sufficient	Fully sufficient

In this article there was confirmed that PD and PCA methods are sufficient for cluster analysis potential identification. The advantages of these methods consist in their predicative ability and the possibility of extension (for specific cases there can be added some new factors).

References:

[1] PORTER, M. E. *The Competitive Advantage of Nations*. New York: The Free Press, 1990.
[2] PORTER, M. E. *On Competition*. Boston: Harvard Business Review Book, 1998.
[3] BERGMAN, E. M., FESER, E. J. *Industrial and Regional Clusters: Concepts and Comparative Applications*. Virginia: WVU Regional Research Institute, 1999.

[4] SKOKAN, K. *Konkurenceschopnost, inovace a klastry v regionálním rozvoji*. 1. vyd. Ostrava: Repronis, 2004.

[5] *Innovative Clusters: Drivers of National Innovation Systems*: Paris: OECD Proceedings, 2001.

[6] SÖLVEL, O., LINDQVIST, G., KETELS, CH. *The Cluster Initiatives Greenbook.* 2006.

[7] ANDERSON, T., SERGER, S. S., SÖRVIK, S. *The Cluster Policies Whitebook.* Malmö: IKED, 2004.

[8] FESER, E. J., BERGMANN, E. M. National Industry Cluster Templates: A Framework for Applied Regional Cluster Analysis, *In Regional Studies*, 2000, Vol. 34, No. 1.

[9] LICHTY, R. W., KNUDSEN, K. R. Measuring Regional Economic Base, *In Economic Development Review*, 1999, Vol. 16, No. 3.

[10] PORTER, M. E. Clusters and the new Economics of Competition, *In Harvard Business Review*, 1998, Vol. 76, No. 6.

Contacts addresses:
Ing. Jan Stejskal, Ph.D.
Institute of Economics
Faculty of Economics and Administration
University of Pardubice
Email: jan.stejskal@upce.cz

Ing. Petr Hájek, Ph.D.
Institute of System Engineering and Informatics
Faculty of Economics and Administration
University of Pardubice
Email: petr.hajek@upce.cz